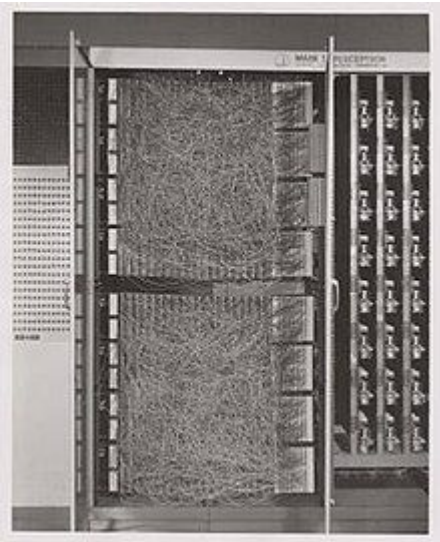
Neural Networks

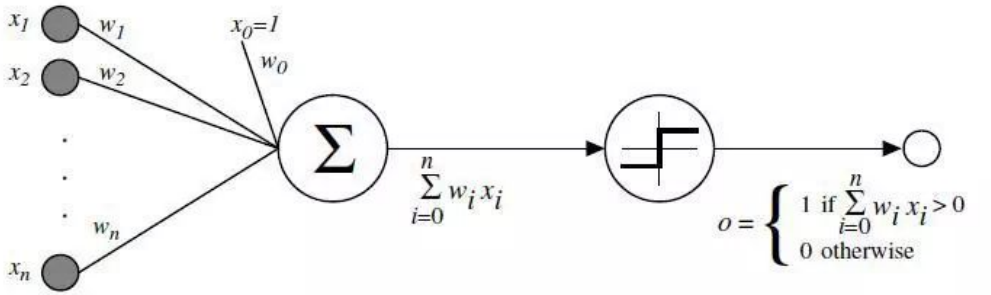
Week 5

History a little bit...

Perceptron - a simple neuron which applies weighted sum to the features X, applies threshold function, and finalizes the binary classification process.

Created in 1958 by Frank Rosenblatt[1]





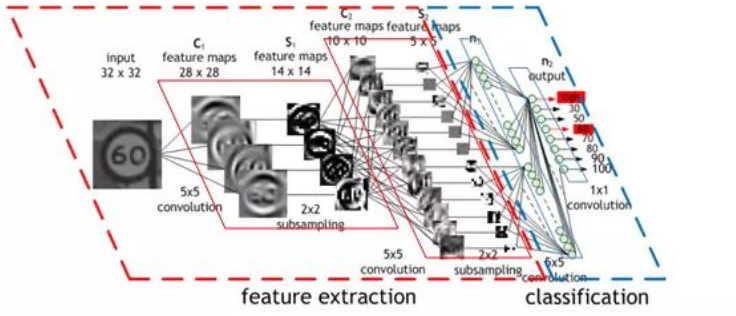
Source: https://en.wikipedia.org/wiki/Perceptron

Starting to improve..

By the 1980s, the blossom happened. A neural network with one hidden layer can now be used as a universal function approximator(1989)[2]

Yann Lecun created a convolutional neural network, which can detect handwritten digits from an 28x28 image[3].

...



Source: https://sites.google.com/site/5kk73gpu2013/assignment/cnn

Going Deeper

ImageNet Large Scale Visual Recognition Challenge[4] has begun in 2010. The challenge is about classify more than 1000 categories.

In 2012, Alexnet[5] has achieved 16% top-5 error rate from 25%.

In 2014, Google submitted Inception V1[6] and got 6.67% top-5 error rate. Follow-up VGG-16[7] accomplished 7.3% top-5 error rate.

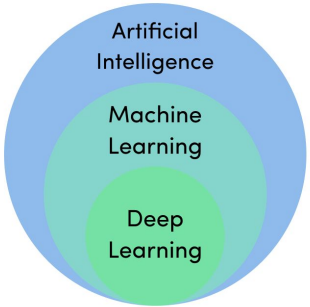
In 2015 ResNet[8], created by Microsoft Research team, got 3.57% error rate.

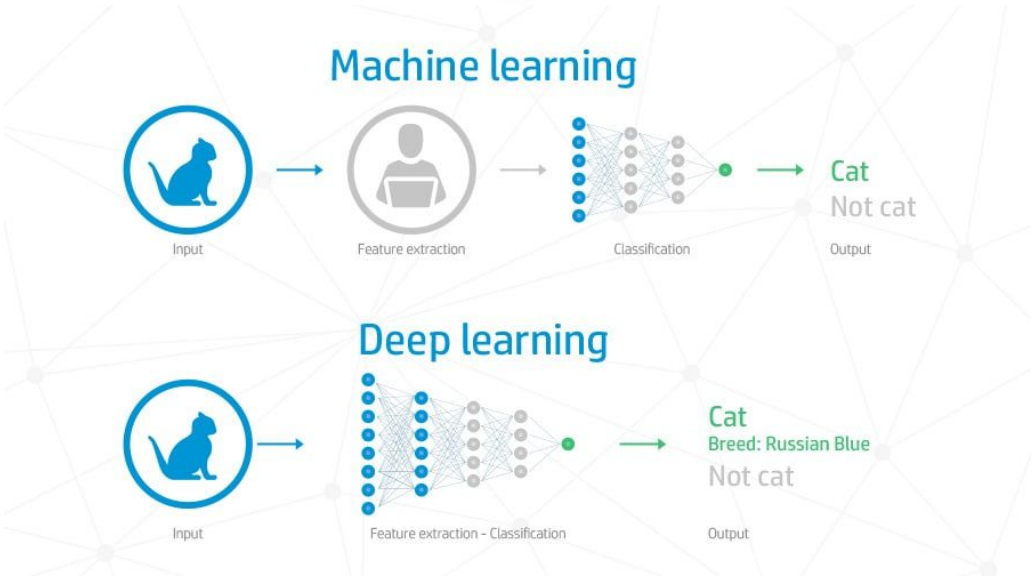
Source TOP: https://www.levity.ai/blog/difference-machine-learning-deep-learning

What is Neural Network

Deep learning is a subpart of Machine Learning

The main difference is, the model can be able to extract relevant features itself, and perform much more better with enough computation power and data.





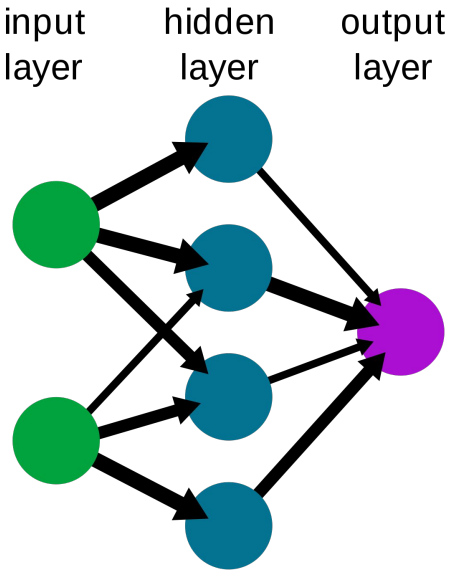
Source bottom: https://www.net-cloud.com/blog/machine-learning-and-deep-learning-101/

A neural network can be simply defined as image on right.

There are inputs which we provide, a hidden layer(s) that process the data, and finally an output layer.

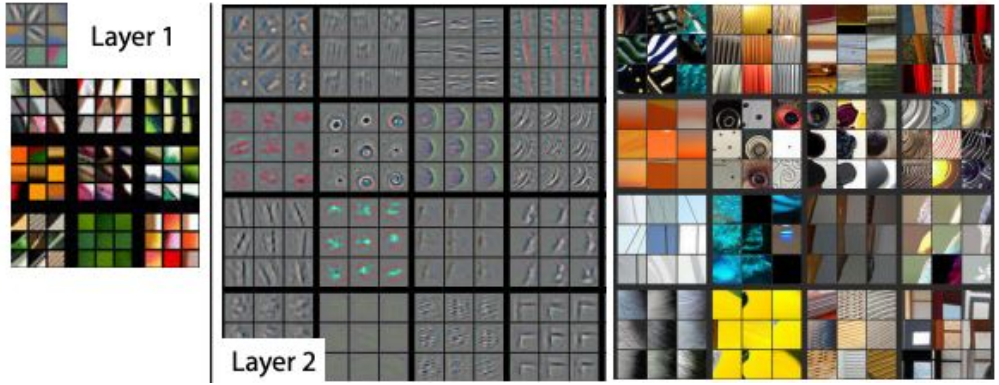
Every neuron inside the hidden layer is calculating a hidden feature that is beneficial for the model.

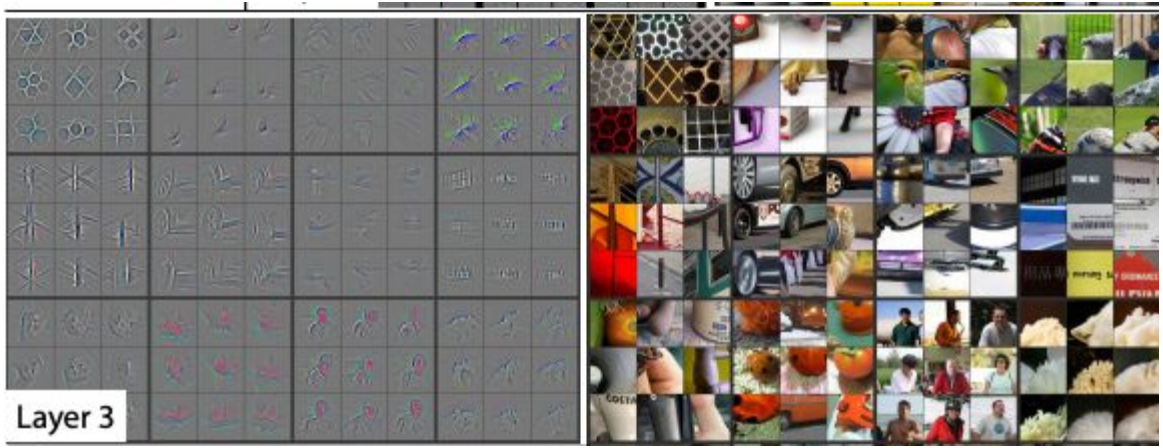
A simple neural network



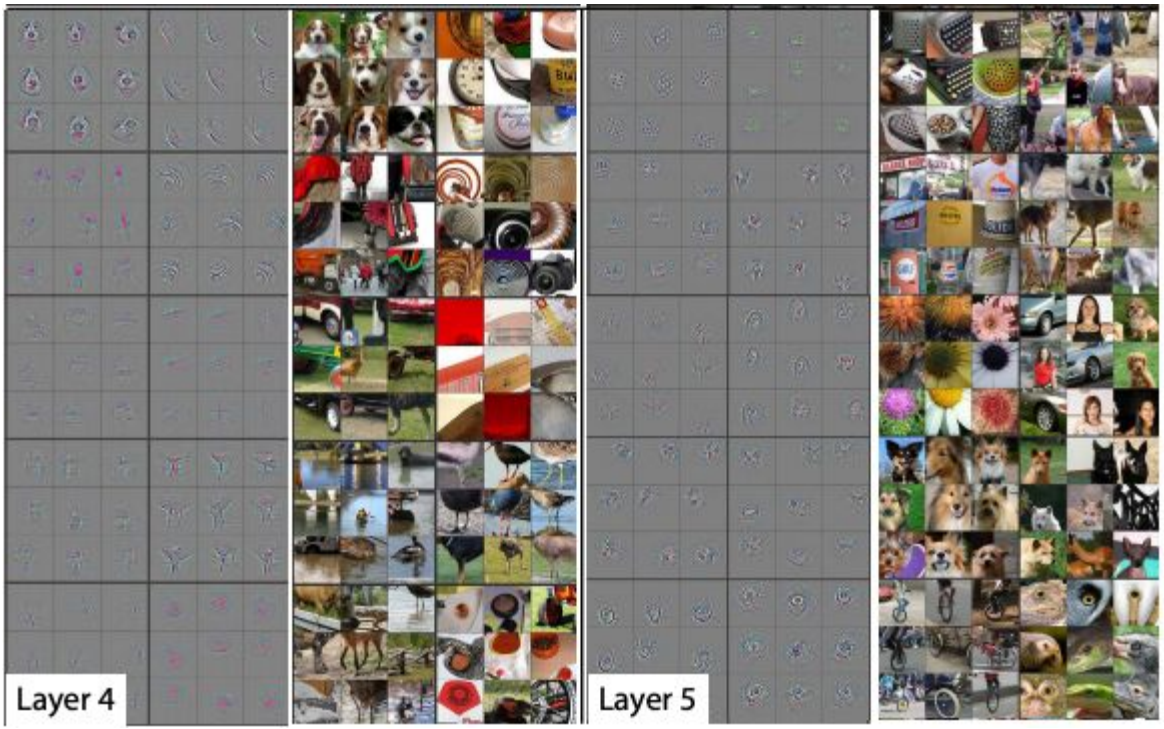
Source: https://en.wikipedia.org/wiki/Neural_network

What do they learn





Zeiler, M. D. and Fergus, R., “Visualizing and Understanding Convolutional Networks”



Zeiler, M. D. and Fergus, R., “Visualizing and Understanding Convolutional Networks”

Forward Pass

Start with the simplest function, which is a linear first order equation $ax + b$

We will compute this function at each neuron, the output of neuron located at first layer will be input for the neuron that is in second layer and similarly for 3rd layers…

But, by computing this straightforward, at the end we get a linear equation.

$$a_n(a_{n-1}(...(a_1X + b_1)...) + b_{n-1}) + b_n$$

$$AX + B$$

Activation Functions

The purpose to non-linearize the outputs of neurons to approximate more complex functions.

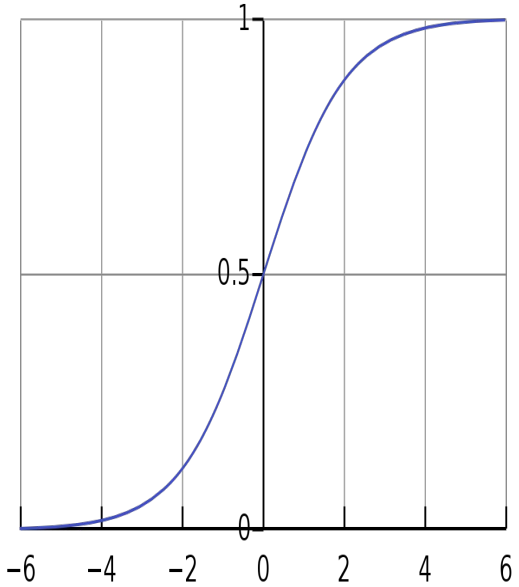
The are also help normalizing the outputs, and sometimes helps the model converge faster.

There are many activation functions, but the most popular ones are:

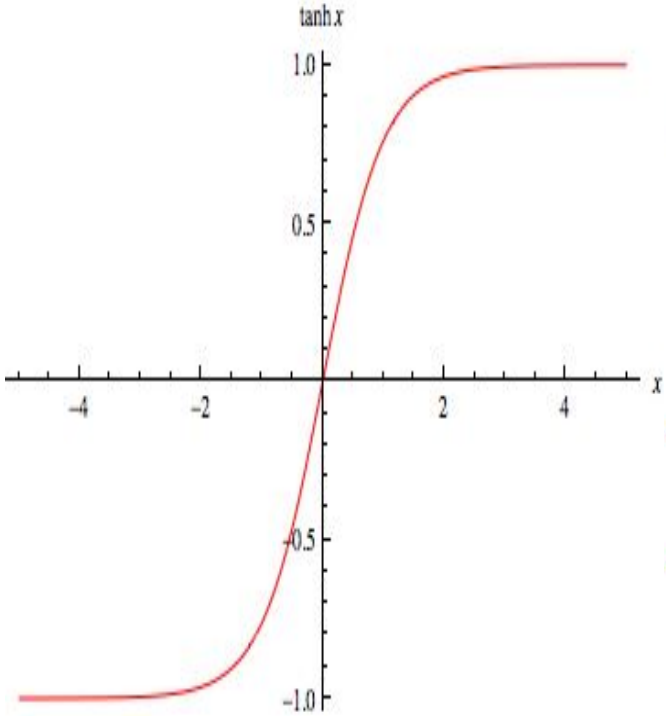
Sigmoid, Hyperbolic Tangent(tanh), ReLU, Leaky ReLU, Softmax,.

The others can be found here: https://en.wikipedia.org/wiki/Activation_function

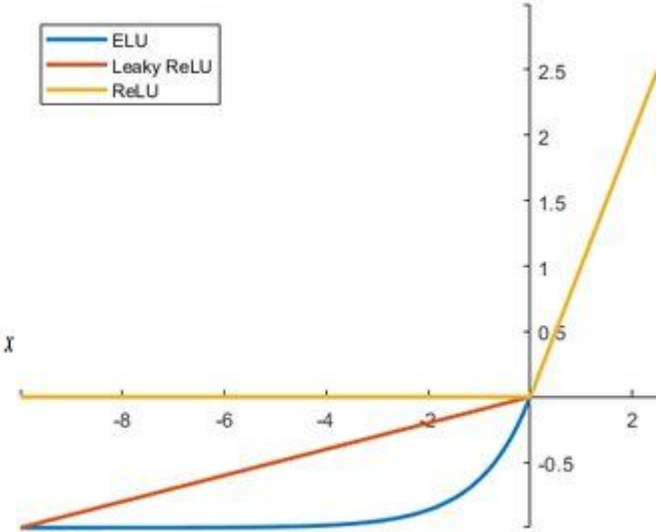
Most heard activation functions



Sigmoid



Tanh



ReLU/Leaky ReLU/ELU

Softmax Classifier

Input pixels, $\mathbf{x}$

Forward propagation

Feedforward output, $\mathbf{y}_i$

cat	dog	horse
5	4	2
4	2	8
4	4	1

Softmax function

Softmax output, $\mathbf{S}(\mathbf{y}_i)$

cat	dog	horse
0.71	0.26	0.04
0.02	0.00	0.98
0.49	0.49	0.02

Shape: (3, 32, 32)

Shape: (3,)

Shape: (3,)

Source: https://ljvmiranda921.github.io/notebook/2017/08/13/softmax-and-the-negative-log-likelihood/

References

[1]Rosenblatt, Frank (1957). "The Perceptron—a perceiving and recognizing automaton"

[2]Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators,1989

[3]Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. 1989. Backpropagation applied to handwritten zip code recognition

[4] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database.

[5] Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.

[6] Szegedy, C., “Going Deeper with Convolutions”, 2014

[7] Simonyan, K. and Zisserman, A., “Very Deep Convolutional Networks for Large-Scale Image Recognition”, 2014

[8] He, K., Zhang, X., Ren, S., and Sun, J., “Deep Residual Learning for Image Recognition”, 2015